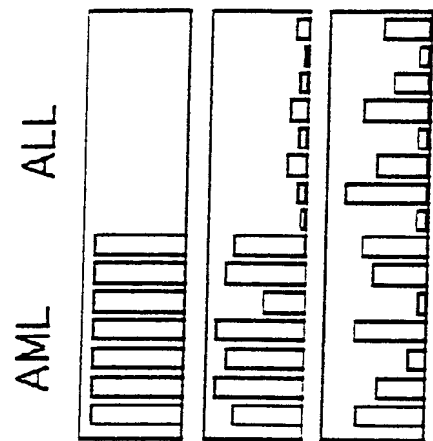


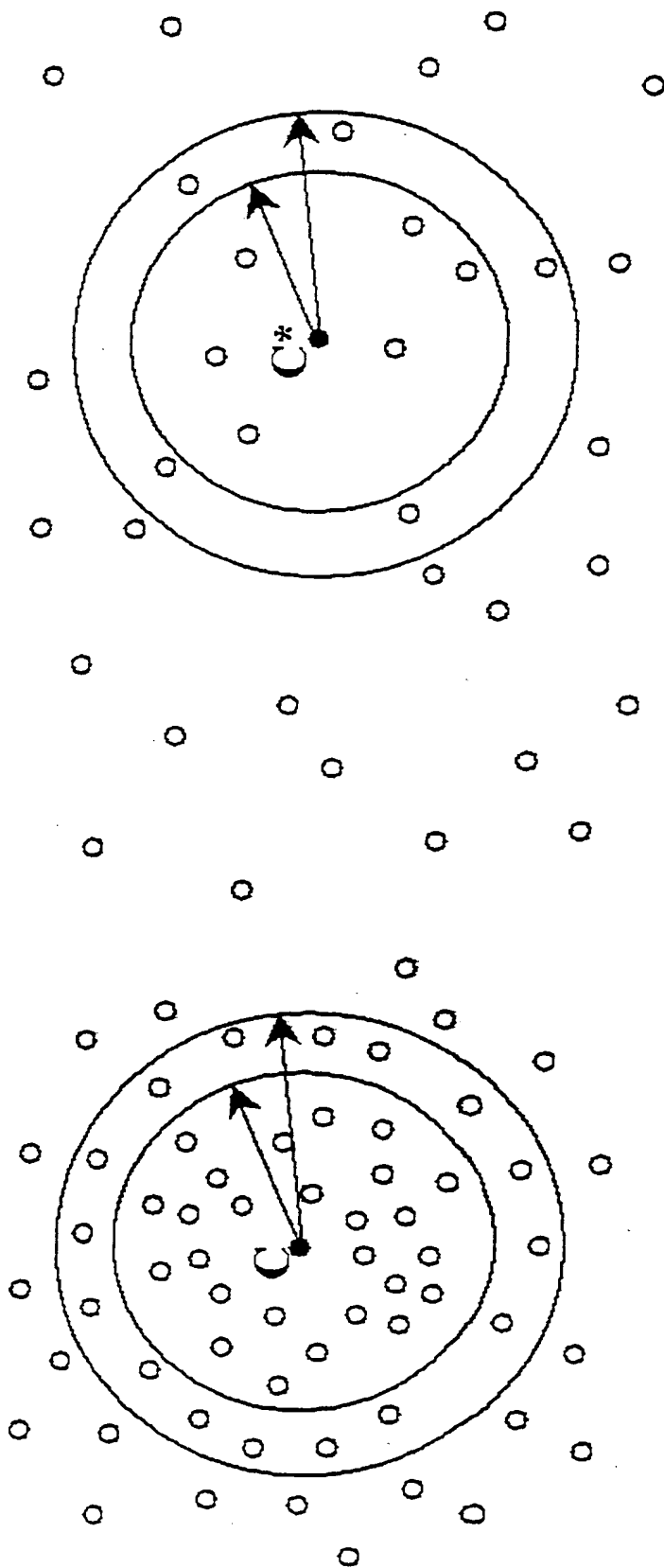
Figure 1A



$$C = (1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0)$$

$$\text{gene}_1 = (e_1, e_2, e_3, \dots, e_{12})$$

$$\text{gene}_2 = (e_1, e_2, e_3, \dots, e_{12})$$



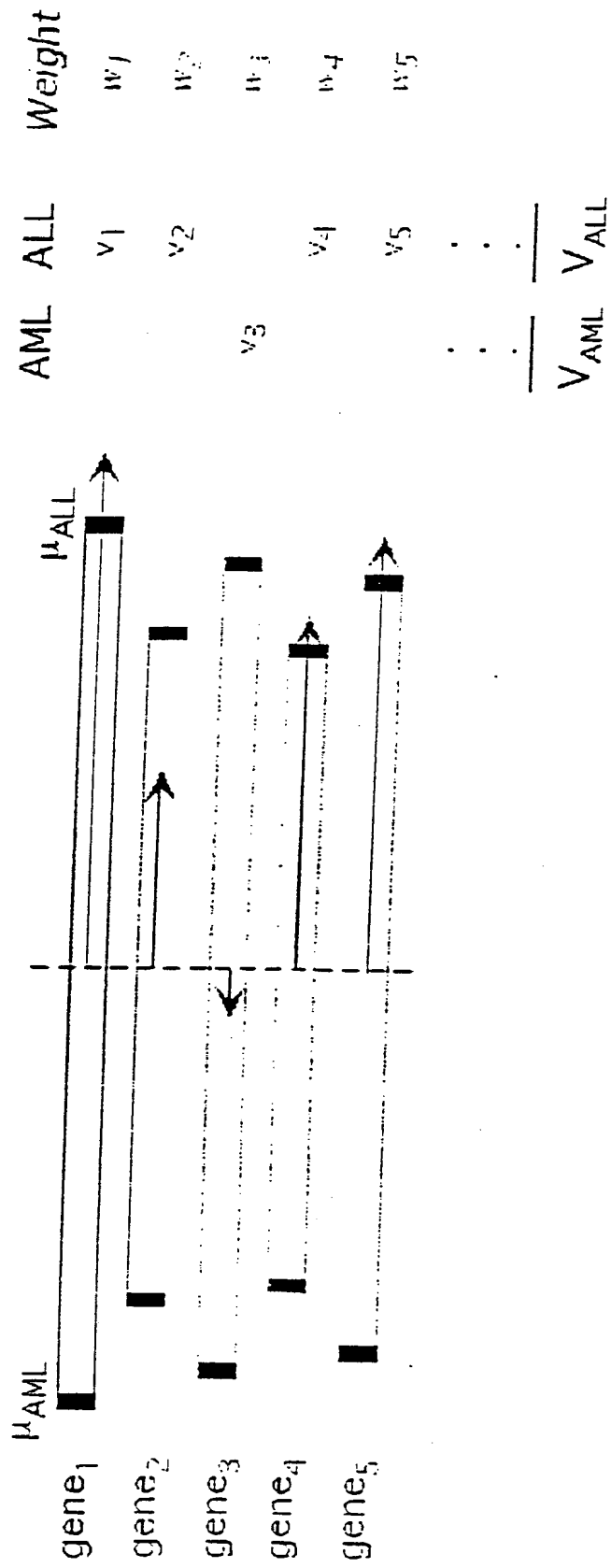


Figure 1C

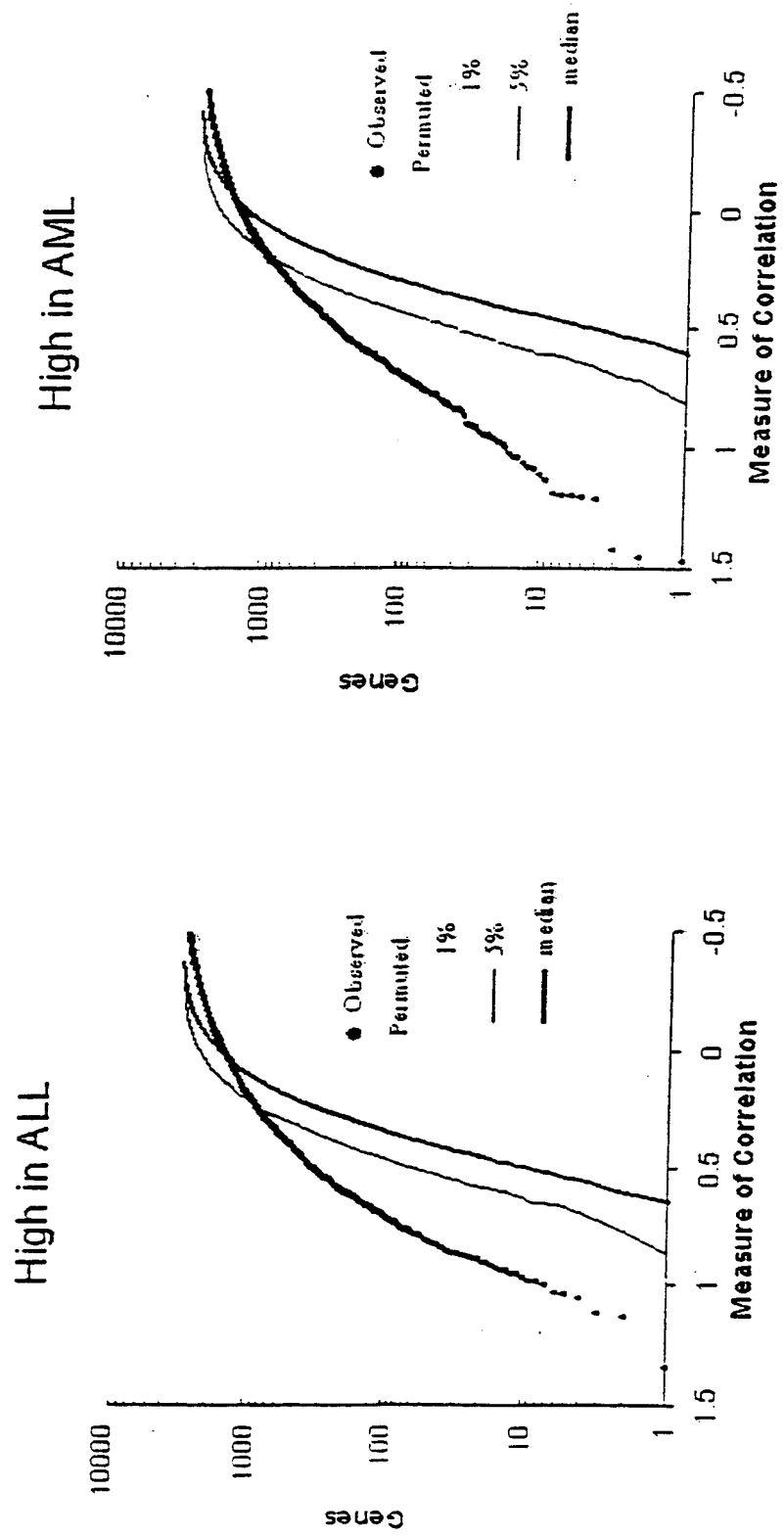


Figure 2

Fig. 3A

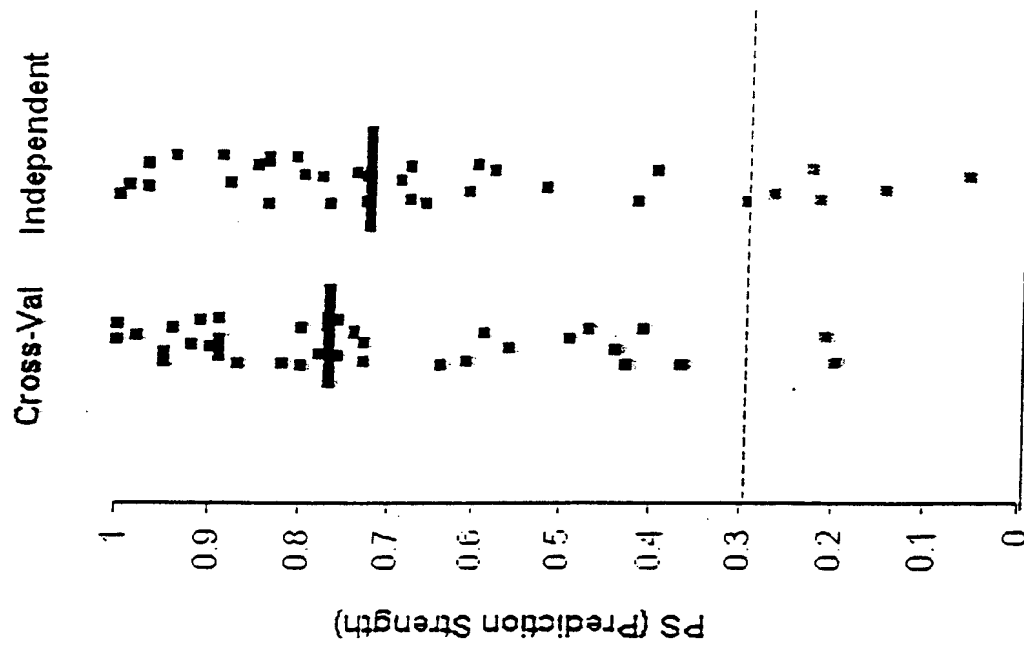
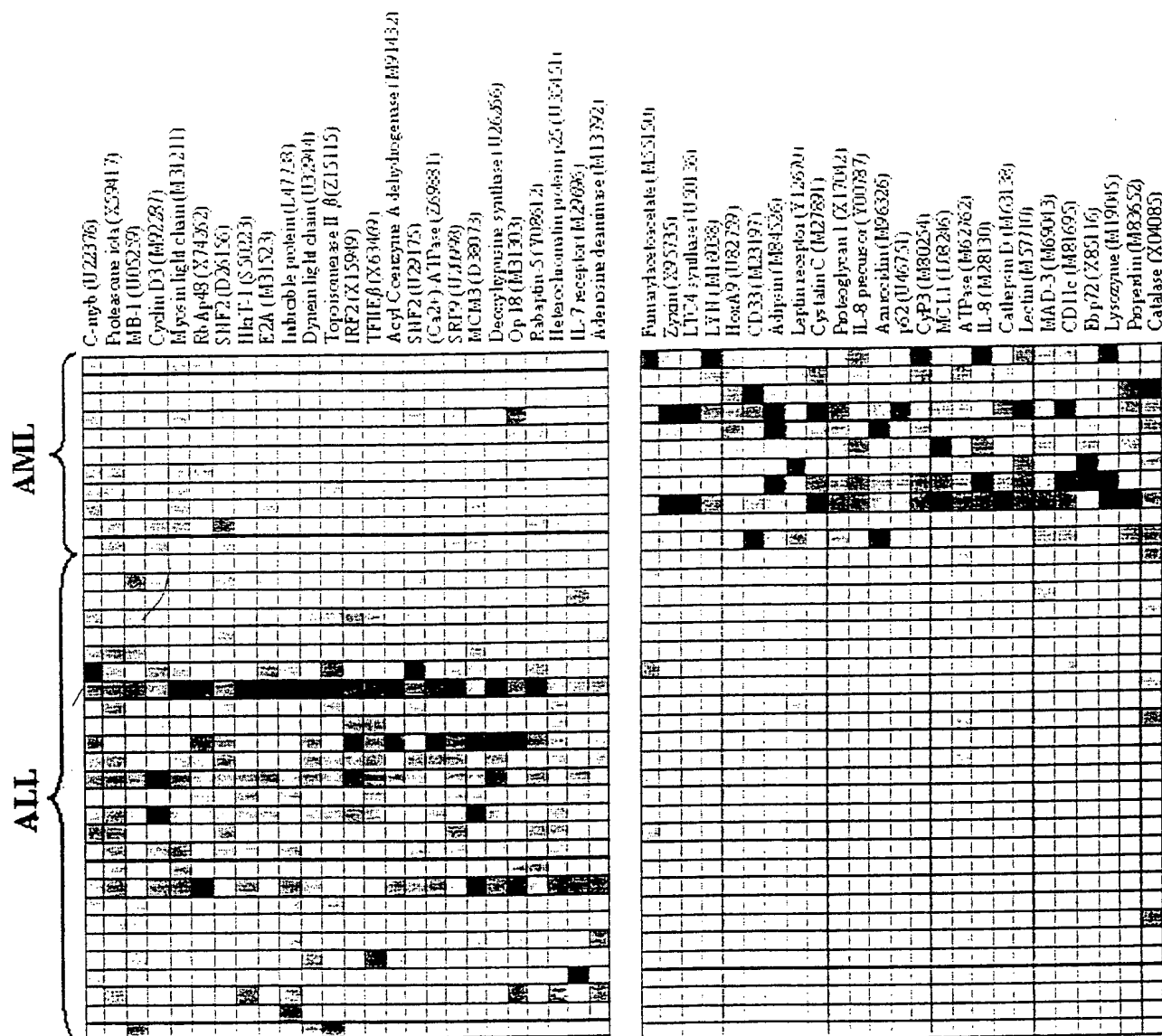


Fig. 3B



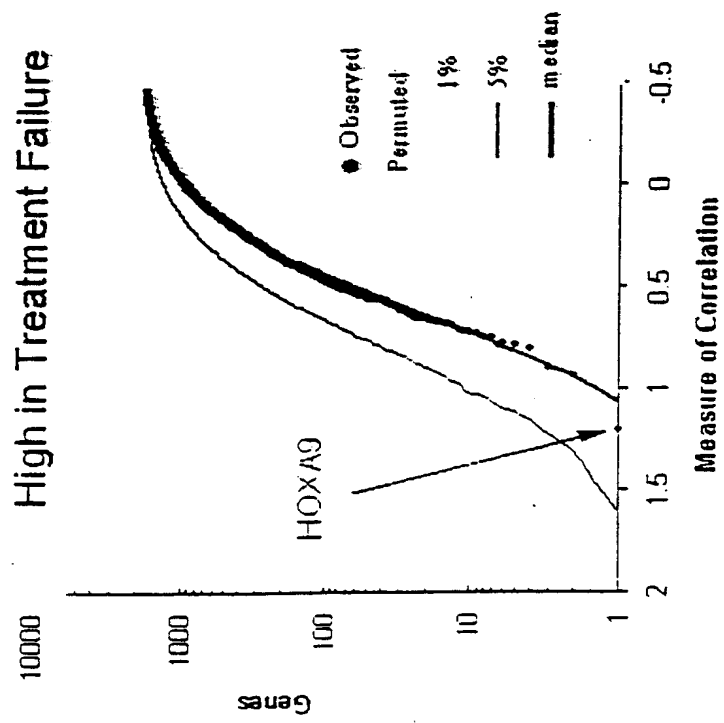
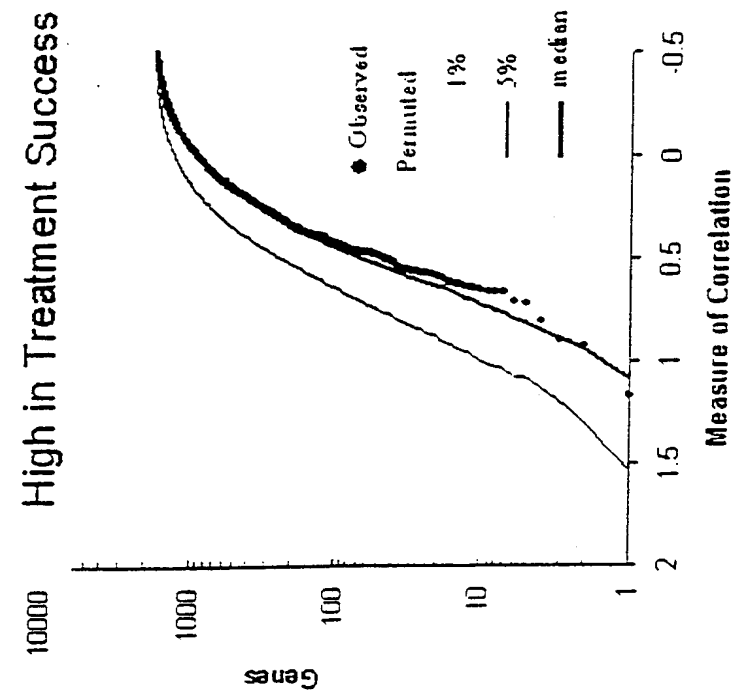


Figure 4

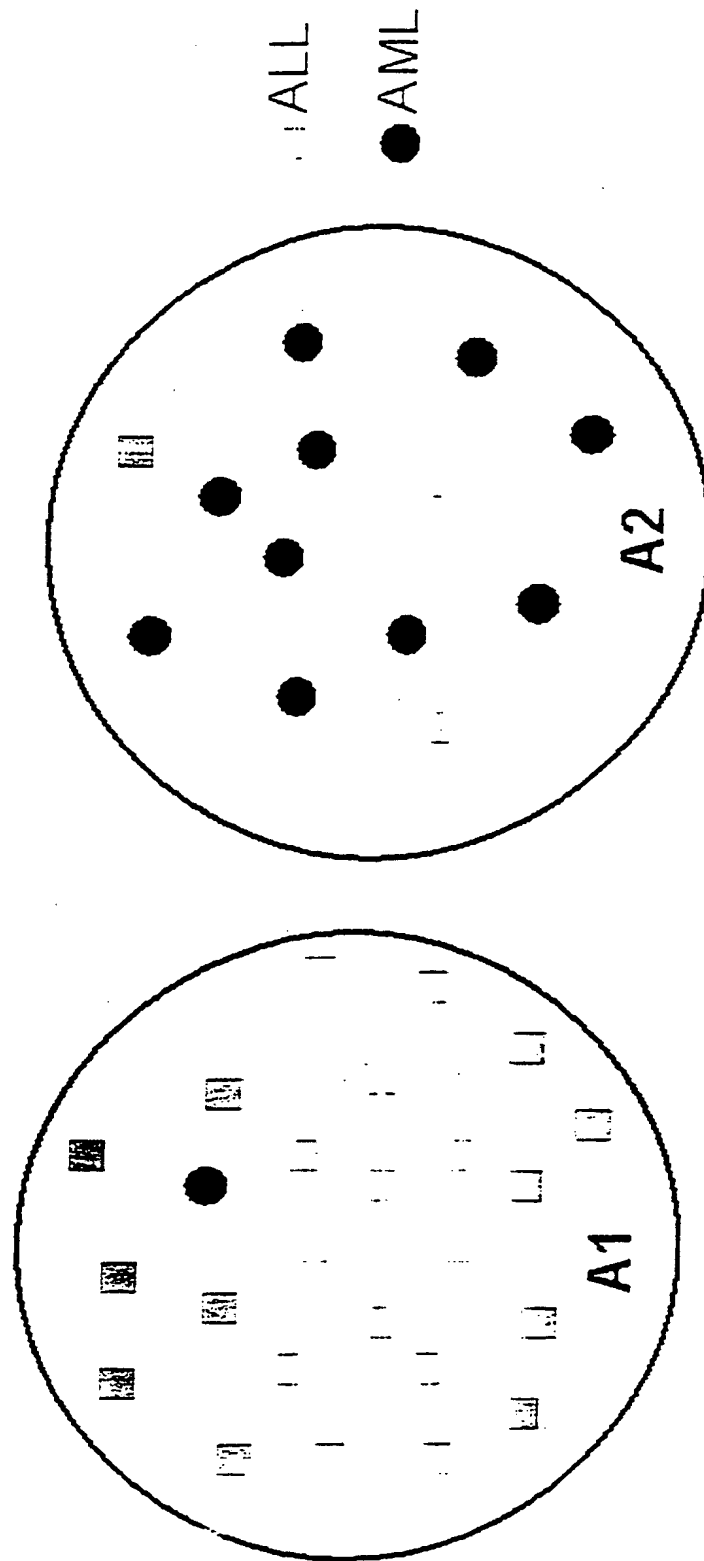


Figure 5A



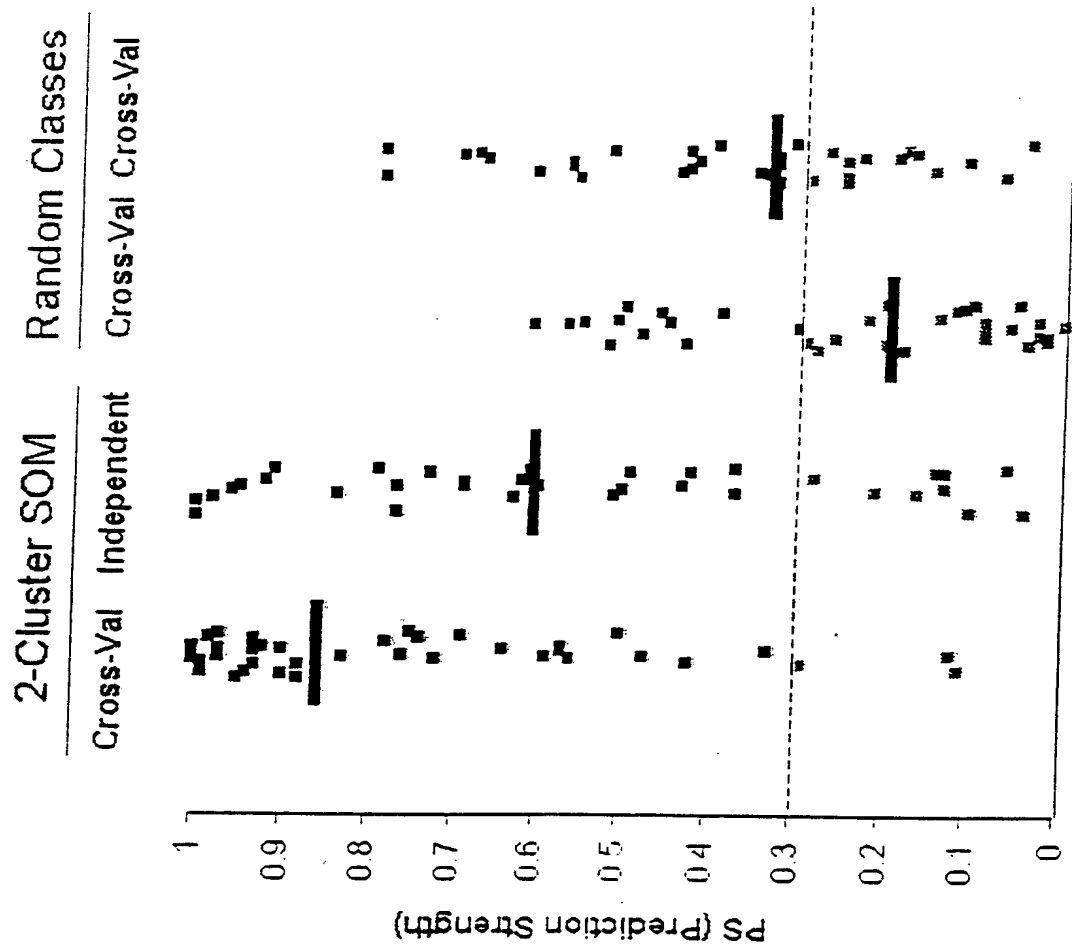


Figure 5B

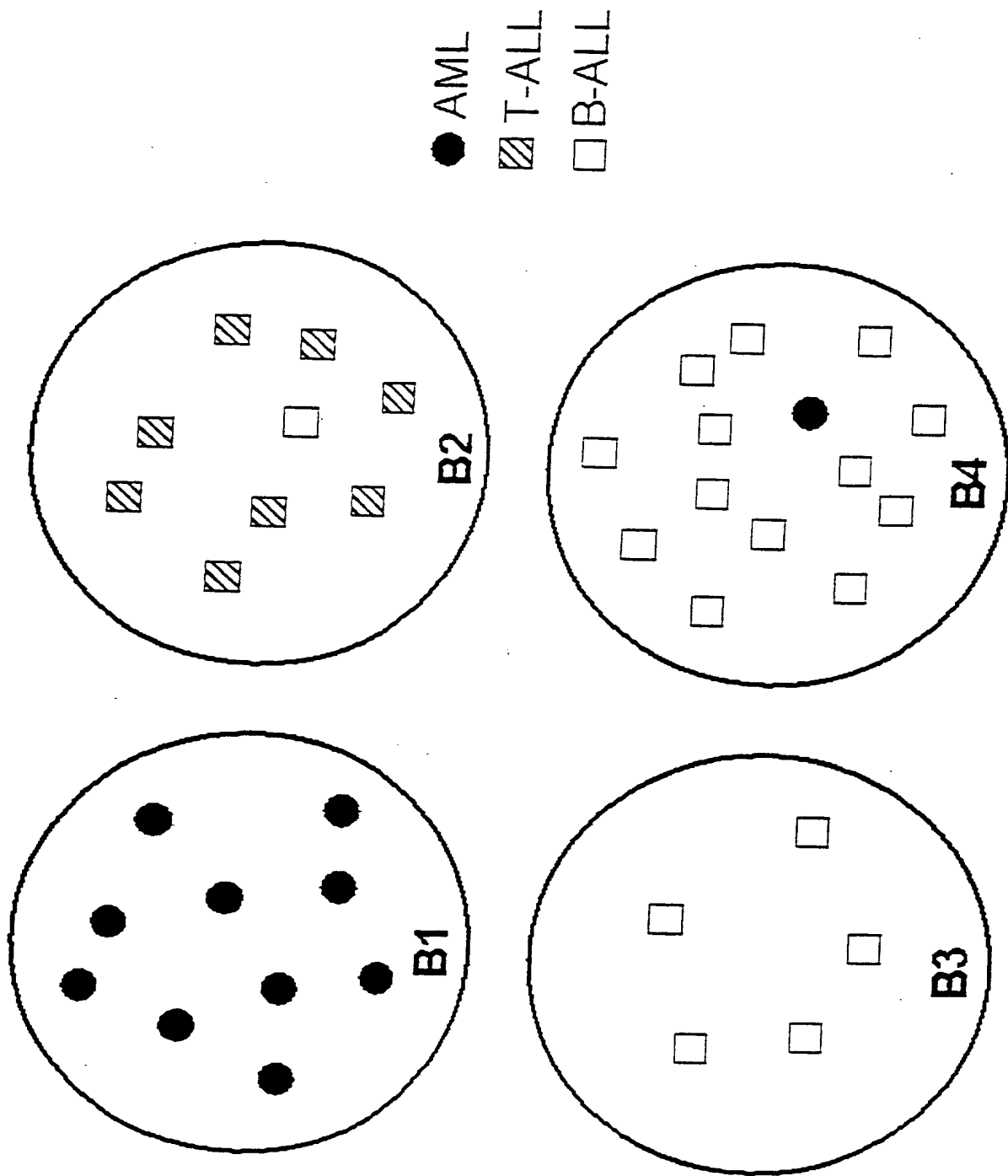


Figure 5C

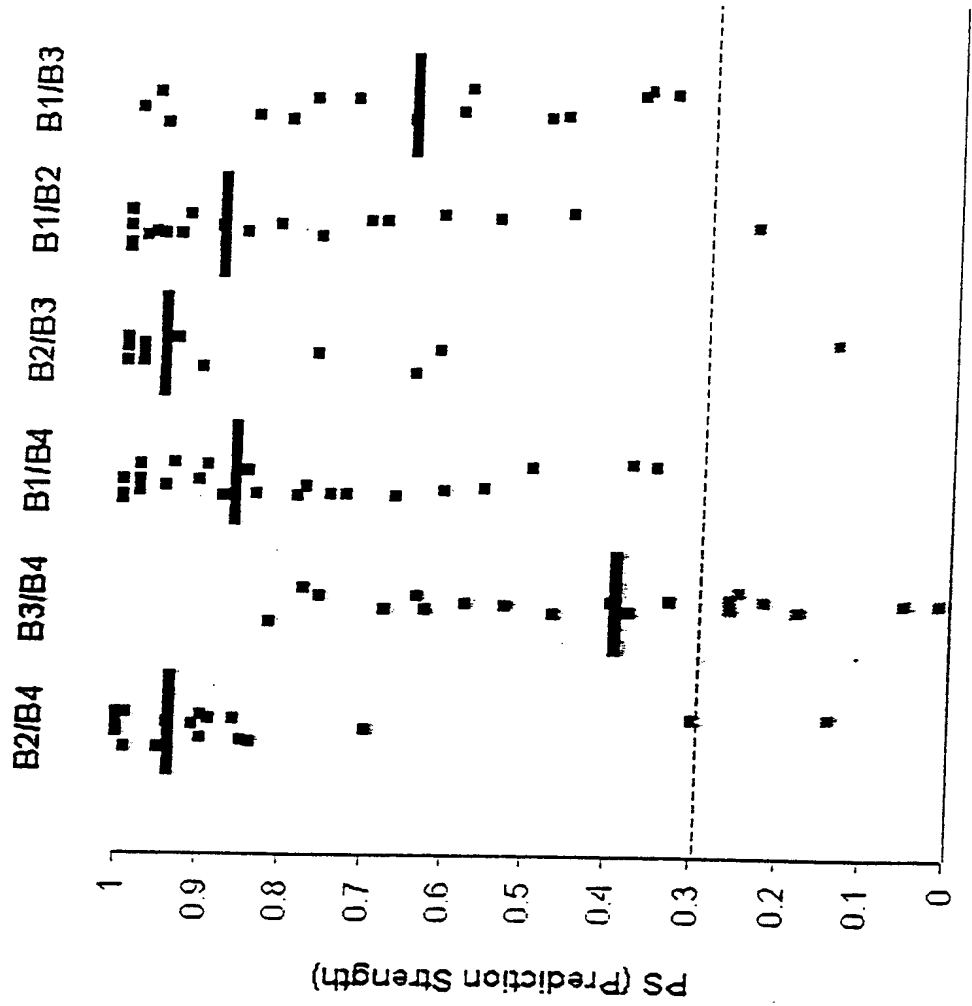


Figure 5D

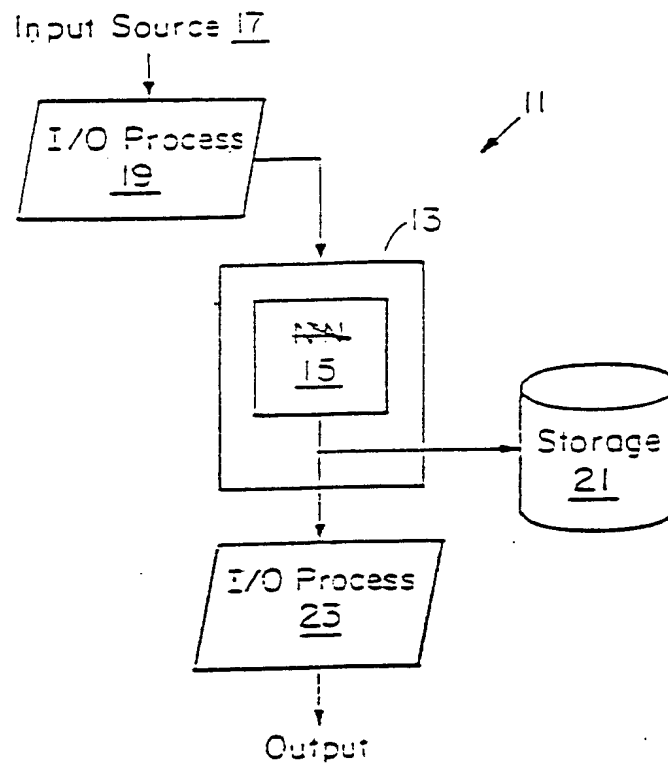


Figure 6

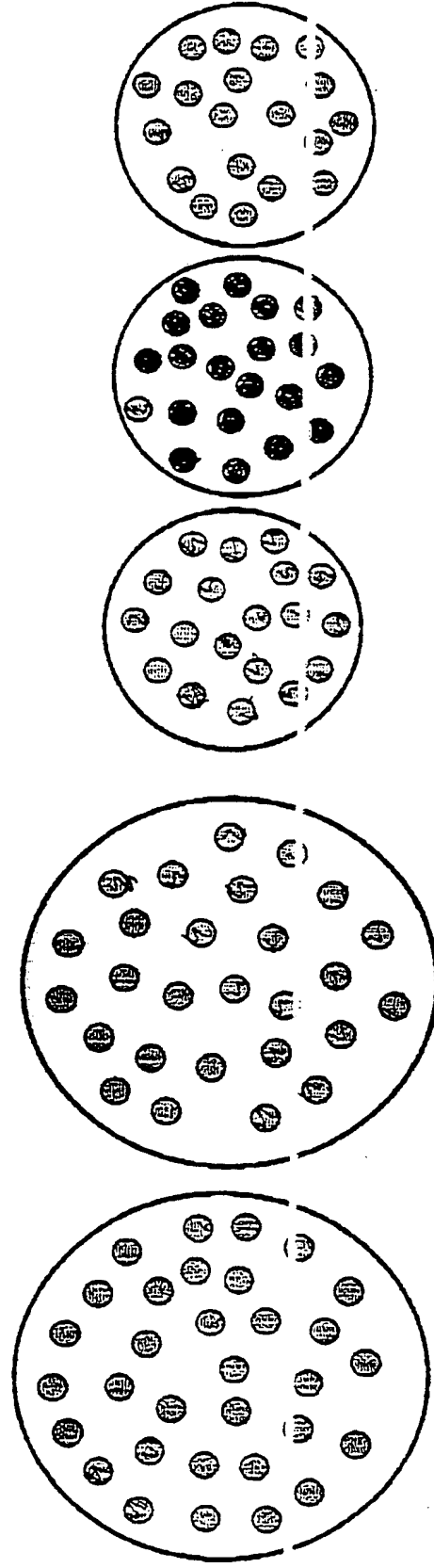
# Example of Self-Organizing Map Class Discovery:

## Lymphoma Large B-Cell and Follicular

Sample collection:

39 large B-Cell + 20 Follicular

Discovered Classes:



●, ⊗ Large B-Cell

⊗ Follicular

Fig. 7

# Example of Self-Organizing Map Class Discovery:

## Brain Glioma and Medulloblastoma

Sample collection:

24 Medulloblastomas + 15 Gliomas

Discovered Classes:

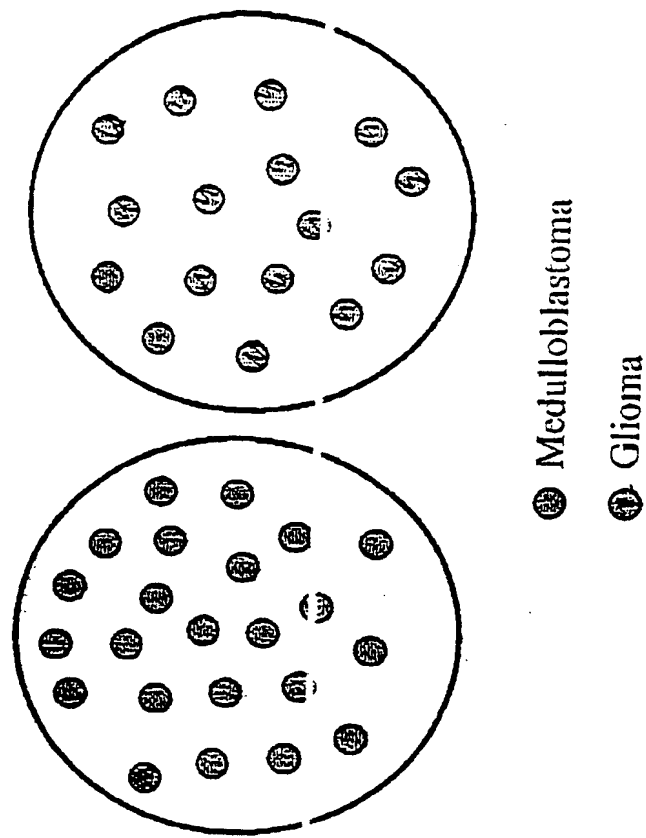


Fig. 8

# Multidimensional Scaling of Leukemia Samples

(431 genes, 15-fold,  $\delta > 1500$ ,  $\text{thres} = 100$ ,  $\text{ceil} = 16,000$ )

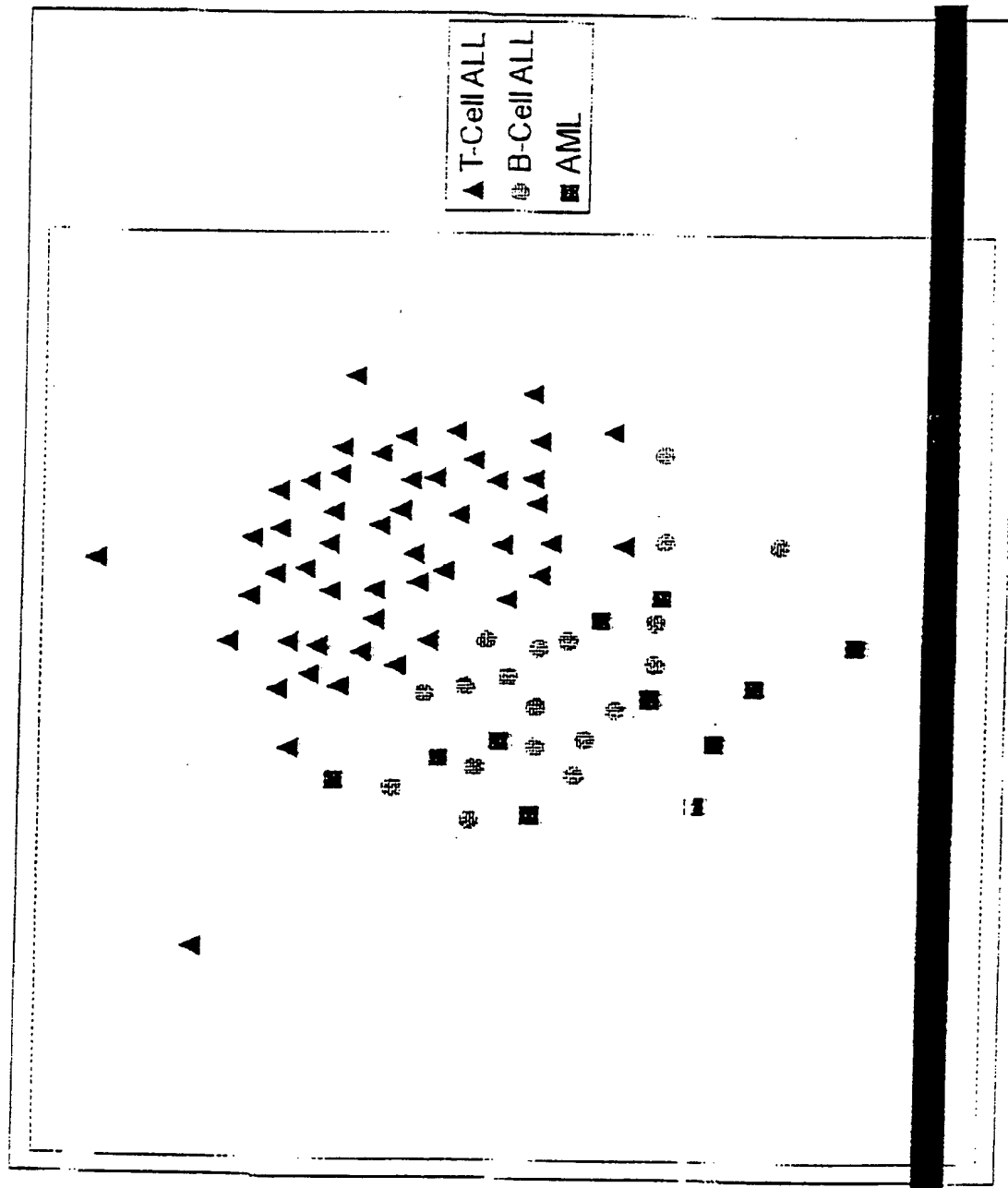


Fig. 9

# Hierarchy of Problems in Molecular Class Prediction




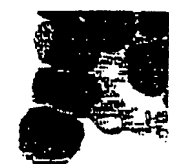




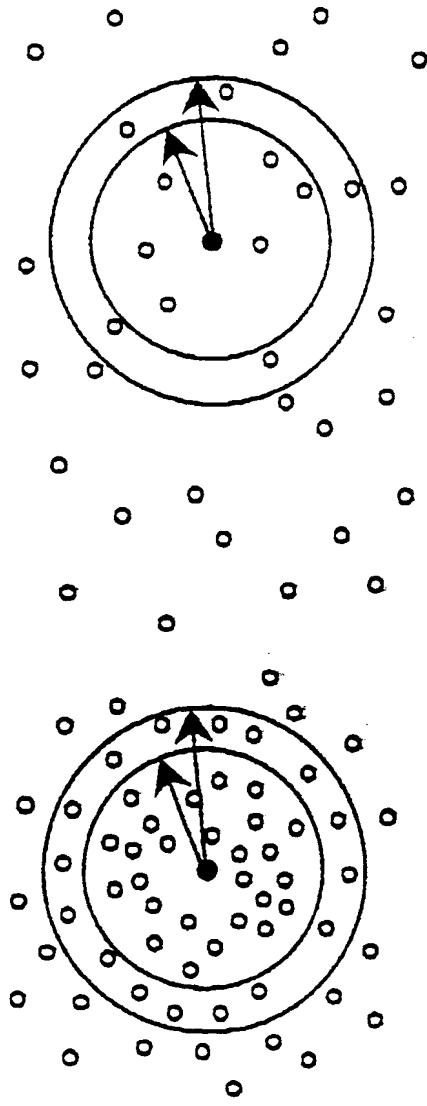
| Problem:                                      | Difficulty: | Gene Markers: | Error: | Example:  |
|---|-------------|---------------|--------|---|
| I. Tissue or Cell Type<br>Normal vs. Abnormal | Low         | ~1000-2000    | ~0%    |  <br>Normal vs. Renal Carcinoma |
| II. Morphological Type                        | Low-medium  | ~200-500      | ~0-5%  |  <br>Leukemia ALL vs. AML       |
| III. Morphological Subtype                    | Medium-high | ~50-100       | ~0-15% |  <br>ALL B- vs. T-Cell      |
| IV. Treatment Outcome<br>Drug Sensitivity     | High        | ~1-20         | ~5-50% |  <br>AML Treatment Outcome  |

Fig. 10



# Neighborhood Analysis: Assessing Statistical Significance of

## Gene-Class Correlations



Class Pattern Neighborhood

Permuted Pattern Neighborhood

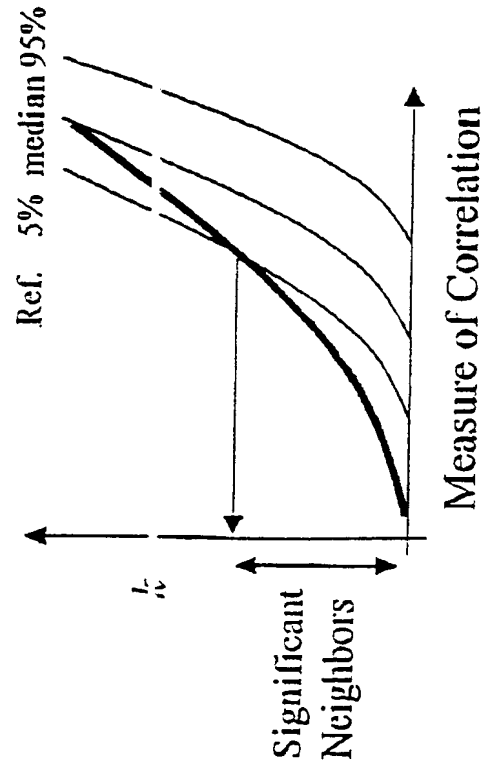
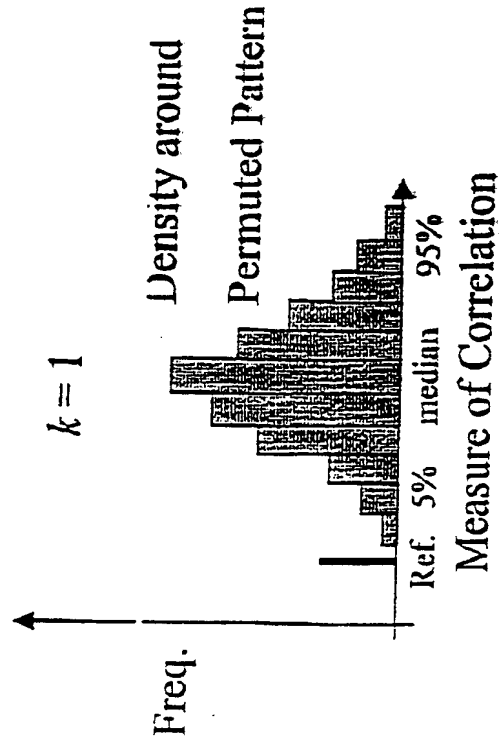


Fig. 11

## Class Prediction Results

| Problem Type | Biological System | Problem Description        | Number of samples | Number of errors | Number of no calls | Number of errors (all calls) | Number of generated |
|--------------|-------------------|----------------------------|-------------------|------------------|--------------------|------------------------------|---------------------|
| I            | Parh              | Normal vs. Cystic Fibrosis | 12                | 0 (0%)           | 0                  | 0 (0%)                       | >100                |
| I            | Leukemia          | ALL vs. AML                | 85                | 0 (0%)           | 2                  | 0 (0%)                       | 700                 |
| I            | Leukemia          | ALL B vs. T Cell           | 33                | 0 (0%)           | 1                  | 1 (3%)                       | 200                 |
| ML           | Leukemia          | Treatment Outcome          | 15                | 2 (13%)          | 0                  | 2 (13%)                      | 1                   |

Fig. 12